

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**  
**ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ**

**Кафедра многопроцессорных систем и сетей**

Аннотация к дипломной работе

**«РАСПРЕДЕЛЕННАЯ ОБРАБОТКА БОЛЬШИХ ДАННЫХ С  
ИСПОЛЬЗОВАНИЕМ ПАРАДИГМЫ MAP/REDUCE»**

Богдан Юлия Сергеевна

Научный руководитель – доцент кафедры МСС, кандидат физ.-мат. наук  
Соболева Т.В.

## РЕФЕРАТ

Дипломная работа, 44 страницы, 22 рисунка, 14 источников.

*Ключевые слова:* BIG DATA, РАСПРЕДЕЛЕННАЯ ОБРАБОКА ДАННЫХ, АНАЛИЗ ДАННЫХ, ПАРАДИГМА MAP/REDUCE, ПЛАТФОРМА HADOOP.

*Объект исследования* – распределенная обработка больших данных.

*Цель работы* – изучение методов обработки больших объемов данных, применение программной модели MapReduce для разработки приложения для распределенной обработки больших массивов метеорологических данных.

*Методы исследования* – изучение технологий обработки больших массивов данных, изучение документации и литературы по рассматриваемым технологиям, анализ реализаций изученных методов.

*Результатом* является изучение и применение парадигмы MapReduce для обработки больших объемов данных, обзор и изучение экосистемы Hadoop, реализация приложения для распределенной обработки метеорологических данных на основе платформы Hadoop, сравнительный анализ производительности при применении различных реализаций рассматриваемой парадигмы в разных средах.

*Областью применения* являются системы анализа и обработки больших объемов информации, средства прогнозирования.

## **ABSTRACT**

Graduate thesis, 44 pages, 22 pictures, 14 sources.

*Keywords:* BIG DATA, DISTRIBUTED DATA PROCESSING, DATA ANALYSIS, MAP/REDUCE PARADIGM, HADOOP FRAMEWORK.

*Object of research* – distributed processing of big data.

*Objective* – the study of methods of processing large amount of data, the use of MapReduce programming model to develop applications for distributed processing of meteorological data.

*Research methods* – the study of the technologies of processing large amount of data, review of documentation and literature on the considered technology, the analysis of the studied methods implementations.

*Results* are the study and application of MapReduce for processing large amounts of data, review and study of the Hadoop ecosystem, the implementation of applications for distributed processing of meteorological data based on the Hadoop platform, comparative performance analysis using different implementations of the paradigms in different environments.

*Scope* is system analysis and processing of large amount of information and forecasting tools.